MASTER COPY KEEP THIS COPY FOR REPRODUCTION PURPOSES

AD-A244 814

ATION PAGE

Form Approved OMB No. 0704-0188

average 1 hour per response, including the time for reviewing instructions, searching existing data sources, reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this is burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson he Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

PORT DATE: mber 30, 1991	3. REPORT TYPE AND DATES COVERED
4. TITLE AND SUBTITLE: Computational techniques for probabilistic inference	5. FUNDING NUMBERS: DAAL03-88-K-0183
6. AUTHOR(S): Edward H. Shortliffe and Gregory F. Cooper	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES): Stanford University, Stanford, California	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES): U. S. Army Research Office, P. O. Box 12211, Research Triangle Park, NC 27709-2211	10. SPONSORING/MONITORING AGENCY REPORT NUMBER: ARO 25514.27-EL

(2)

11. SUPPLEMENTARY NOTES

The view, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.

12a. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution unlimited

92-00453

13. ABSTRACT (Maximum 200 words)

The objectives of this research project were to develop pragmatic and theoretically sound methods for the computation of probabilistic information within expert systems. We explored the use of Bayesian belief networks as a probabilistic representation. We implemented and evaluated several previously described belief-network inference algorithms that perform exact inference, as well as developing a hybrid algorithm and a new algorithm. Our conclusion is that no single algorithm is best for all inference problems. Moreover, our analysis revealed that the belief-network inference problem is NP-hard. Thus, it is unlikely we can develop an exact algorithm that is uniformly efficient (polynomial time) across all networks and inference problems. This led us to investigate special-case and approximation algorithms, as well as methods for controlling multiple algorithms in solving a single inference problem. Our investigation indicates that moderately complex expert systems based on belief networks can be constructed using these current methods. The development of improved methods for controlling the application of multiple inference algorithms is likely to allow tractable inference in increasingly complex expert systems based on belief networks. The construction of complex belief networks also presents significant challenges. We developed automated and semi-automated knowledge-acquisition techniques which show significant promise in preliminary tests.

14. SUBJECT TERMS: Bayesian belief networks, probabilistic expert systems, probabilistic inference, knowledge acquisition, machine learning	15. NUMBER OF PAGES: 11
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

Computational Techniques for Probabilistic Inference

Statement of the Problem Studied

Decision making typically is replete with uncertainty. In particular, there is uncertainty due to incomplete and inexact models, and uncertainty secondary to incomplete and erroneous data. Therefore, in general, it is important that computer systems that assist in decision making be capable of representing and reasoning with uncertainty. In this project we have explored the use of probability theory as a representation of uncertainty in diagnostic systems. There are several advantages to using a probabilistic representation, including that it (1) is mathematically well-defined and has been studied extensively, (2) provides a common, well-established language for communicating uncertainty, (3) allows the combination of subjective probabilities from medical experts with statistics gathered from databases, and (4) can be naturally extended to a decision-theoretic system that recommends actions to take. Nonetheless, there are potential problems associated with using a probabilistic representation. Key challenges include developing tractable methods for knowledge acquisition and probabilistic inference. During the last three years we have addressed these two problems using the belief-network representation. Belief networks provide a graphical representation for efficiently and intuitively specifying the probabilistic dependencies among domain variables.[1]

Summary of the Results

In this section, we summarize our results on belief-network inference and acquisition.

Probabilistic inference

Studying and extending cutset conditioning

When we began work on this ARO project, Pearl had only recently described a new belief-network inference algorithm based on message passing and cutset-conditioning (call it the CC algorithm). We chose to initiate our study of belief-network inference algorithms by implementing the CC algorithm; to our knowledge, we were the first to implement the algorithm in its general form. In the process, we worked out many of the technical details that previously were unspecified [22]. In particular, we examined cutset conditioning on multiply-connected networks. We proved that finding a minimal cutset is NP-hard, and we developed and evaluated a heuristic for finding small cutsets [19].

[1] For a detailed discussion of the belief-network representation, see J. Pearl, *Probabilistic Reasoning in Intelligent Systems* (Morgan Kaufmann, San Mateo, CA, 1988).

A-1

An evaluation and combination of two previous algorithms

In 1988 Lauritzen and Spiegelhalter published a new algorithm for belief-network inference based on clique-tree propagation, which we implemented (call it the CTP algorithm). In [21] we analyze some of the strengths and weaknesses of the CC and the CTP algorithms. We also empirically evaluated both algorithms on a 37-node network called ALARM and found that the CTP algorithm performs probabilistic inference significantly faster than the CC algorithm; in [1] we discuss the reasons why. The insights gained from implementing and evaluating these two algorithms led us to develop a hybrid algorithm that combines their strengths [21]. In [20] we show empirically that the hybrid algorithm decreases inference time when applied to the Pathfinder knowledge base.

A new inference algorithm based on recursive decomposition

Although the hybrid algorithm performs well in many cases, there are cases when it does not. We developed a new belief-network inference algorithm called recursive decomposition (RD), which handles some of these cases efficiently [10]. The basic idea of recursive decomposition is to reduce a belief-network inference problem by dividing it into a set of simpler problems. In one form, recursive decomposition bisects a network B into subnetworks B_1 and B_2, using a set of nodes S, called the vertex separator set. The decomposition procedure is applied recursively to successively smaller networks until the resulting networks are so small that their solutions are immediate. The solutions to the simpler problems are combined to solve the original problem. There are belief networks for which some types of inferences are exponentially faster using recursive decomposition than CC or CTP. Conversely, there are cases when CC or CTP is more efficient than RD. Thus, RD, CC, and CTP are complementary in that each has its relative strengths and weaknesses.

Complexity analysis of belief-network inference

In [8] we show that probabilistic inference on belief networks is NP-hard. Thus, it is not surprising that researchers have been unable to find a general, exact algorithm that has a polynomial time complexity in the worst case. Unfortunately, in practice there are large, complex belief networks for which general, exact algorithms such as CC, CTP, and RD perform inference too slowly [16, 18]. This led us to explore special-case algorithms and approximation algorithms, which we now describe in turn.

Special-case algorithms

We can decrease the expected inference time by storing (precomputing) the answers to inference problems that are likely to occur. In [13], we discuss methods for applying this technique to belief-network inference. For the ALARM belief network [1], the precomputation led to a two-fold decrease in the expected time to answer a probabilistic query [13]. We consider precomputing to be a special-case technique, because the answer to a query may not always be precomputed due to limitations of storage and time available for precomputation.

Approximation algorithms

Likelihood weighting (LW) is a Monte Carlo simulation method for belief-network inference that was reported in 1989 by Shachter & Peot and by Fung & Chang. We applied LW to the problem of inference on QMR-DT. In particular, for a set of findings, we were interested in determining the posterior probability of each of 600 potential causes of the findings. We assumed that multiple causes are possible. In [18] we describe the QMR-DT model in detail. We compared the QMR-DT model to the QMR model from which it was derived. QMR is a well-known medical diagnostic system developed at the University of Pittsburgh over the last two decades. Previous evaluations of QMR have demonstrated that it performs well in practice on difficult cases when compared to clinicians. QMR uses a tailored, ad hoc scoring scheme for ranking diagnoses. Our evaluation of QMR-DT, using LW as an inference algorithm, shows that its diagnostic accuracy is comparable to that of QMR [16]. This result is encouraging, since QMR-DT did not have access to some forms of knowledge that were available to QMR; thus, we might expect QMR-DT's performance to improve further, after we extend its model. Additional testing will be necessary to investigate the impact of such extensions. Regarding computation time, our QMR-DT simulations required about 90 minutes per case on a Macintosh IIci. In [17] we report our analysis of the specific extensions to the basic LW algorithm that led to the most rapid convergence of the posterior probabilities. Although 90 minutes is too slow to be very practical, there currently are workstations that are several-fold faster than the Macintosh IIci; furthermore, in the next decade we almost certainly will see further significant increases in hardware speed. In addition, the LW algorithm is readily amenable to parallelization. On a parallel computer, we can obtain a decrease in inference time for this task that is nearly proportional to the number of processors [16]. Thus, even for large belief networks like QMR-DT, LW seems to hold significant promise as a practical inference method.

In 1987 Pearl published an algorithm for Monte Carlo simulation of belief networks based on Markov state transitions (called it the MST algorithm). Both MST and LW lack a theory of convergence, which makes it difficult to know how long to run the simulations. In one belief network, we observed during repeated simulations that the MST algorithm got trapped in a portion of the Markov state space and did not converge; in [6] we analyze why such traps occur and we offer some suggestions for avoiding traps. We also derived a theoretical analysis of the worst-case expected convergence of the MST algorithm [4], and in [24] we prove a tight worst-case bound. We developed a derivative of MST called BN-RAS, and in [2] we evaluate the convergence of BN-RAS on two belief networks. The results show that our worst-case theoretical analysis is conservative relative to the empirical convergence that we observed. In [5] we extend the convergence-analysis techniques to logic sampling, which is another simulation method that is closely related to LW.

So far, we have described methods for finding exact, point probabilities or for finding estimates of probabilities using simulation. A third approach that we have explored is to relax our goal to one of determining upper and lower posterior probabilities. In [15] we show that usefully tight bounds are derivable in less time than is required to derive point probabilities in the ALARM [1] domain. In [26] we further explore the derivation of bounds and their practical significance.

Regarding belief-network inference, we have focused most of our efforts on efficiently computing posterior probabilities of the form $P(X \mid Y)$, where X and Y are sets of instantiated variables (i.e., variables with known values). In [7], however, we show how to use algorithms that compute $P(X \mid Y)$ to compute $P(S_1 \mid S_2)$, where S_1 and S_2 are well-formed formulas in propositional logic (propositions).

Controlling probabilistic inference

In [14, 26] we describe our progress in developing decision-theoretic methods for controlling probabilistic inference. In this work we address the question, "How long and with which methods should a computer system deliberate about a probabilistic inference problem before making a recommendation for how to act based on that inference?" In particular, we investigated an approximation algorithm that incrementally tightens bounds on posterior probabilities as more computation time is expended. The critical question is: when are the bounds sufficiently tight for their intended use? The answer to this question depends on a number of factors, including (1) the stakes of the situation at hand, (2) the costs of deliberation, and (3) meta-level knowledge about the expected value of continuing to reason. In the general case, there may be uncertainty about all three of these factors. In [14, 26] we discuss some theoretical principles of belief and action under bounded resources and incomplete inference. We developed techniques that use information about the amount of time required to solve previous complex problems in a domain to determine which techniques to apply in solving current complex problems in that domain. In [26] we describe in detail a graphics-based software system for experimenting with control of probabilistic inference, along with experimental results from its application.

Acquisition of probabilistic models

Computer-assisted acquisition of belief networks from experts

We developed a general-purpose shell called KNET for constructing belief networks using a graphical interface [24, 3[2]]. A knowledge engineer enters a belief network structure by drawing a directed acyclic graph on a monitor using a mouse. The KNET architecture defines a complete separation between the user interface and a belief-network inference-engine subsystem. The inference subsystem contains several of the algorithms discussed in the previous section. A user can select an algorithm to apply in

[2] This paper received first place in the student paper competition at the 1989 Symposium on Computer Applications in Medical Care.

a given case; this capability facilitated our experimentation with several of the inference algorithms discussed in the previous section. We entered four different belief networks using the KNET system. Our experience suggests that a graphical interface such as KNET is useful for entering networks that contain up to several dozen nodes.

The acquisition and application of probabilistic models may be facilitated significantly by having a system that can explain belief-network inference. For example, an expert can use automatic explanations of test-case results as feedback during the belief-network construction process. An explanation system also could provide additional insight about inference results to the end user of a probabilistic expert system. Currently, we are pursuing the development and evaluation of methods that explain the propagation of probabilistic information along pathways in a belief network [23, 27]. Such explanations can guide the process of editing and refining belief-network structures and probabilities.

Computer-based automated generation of probabilistic networks

As stated in the previous section, recent research has led to progress in developing manual methods to improve the efficiency of knowledge acquisition directly from experts. These methods are likely to remain important in domains of small to moderate size in which there are readily available experts. Some domains, however, are large. In others, there are few, if any, experts. Methods for assisting, or in some cases replacing, the manual expert-based methods of knowledge acquisition are needed. We have explored techniques for the automated construction of belief networks. One method involves reducing a large, comprehensive model to a problem-specific model [11]. Another approach involves constructing belief networks from databases.

Databases are becoming increasingly abundant in many areas, including science, engineering, and the military. In each of these areas, there are many potential opportunities for using belief networks to provide assistance in decision making. By using databases to assist in constructing belief networks, we may be able to significantly decrease knowledge acquisition time. Automatically generated networks could be used directly to provide decision-making assistance, or used as a starting point for modification by an expert. In the latter case, the editing of a network may require substantially less time than de novo generation of the network by an expert.

The automated construction of belief networks also can provide insight into the probabilistic dependencies that exist among the domain variables. One application is the automated discovery of dependency relationships. The computer program searches for a belief-network structure that has a high posterior probability given the database, and outputs the structure and its probability. A related task is computer-assisted hypothesis testing: the user enters a hypothesized structure of the dependency relationships among a set of variables and the program calculates the probability of the structure given a database of cases on the variables. These applications clearly have the potential to effect broad areas of discovery and data evaluation.

We have developed two techniques for constructing belief networks from databases. One of them uses an entropy-based approach [12] and the other uses a Bayesian approach [9]. Preliminary results of these two techniques are promising. For example, using the Bayesian approach, we attempted to reconstruct the ALARM belief network [1] from a database of 3,000 cases that we generated earlier using ALARM. Of the 46 arcs in ALARM, the reconstructed network had one arc not in ALARM (a false positive) and it had one arc missing that is in ALARM (a false negative). A subsequent analysis revealed that the missing arc is not strongly supported by the 3,000 cases. The extra arc was added due to the greedy nature of the search algorithm we used. The reconstruction required approximately 5 minutes when running on a Macintosh II computer. In [25] we explore in detail the theory and empirical evaluation of the entropy and Bayesian methods of automated belief-network construction from data. On the basis of our current results and analysis, the Bayesian method appears to be the preferred approach, due to its relative speed, sensitivity, and flexibility.

Summary

The objectives of this research project, as stated in the original proposal, are to develop pragmatic and theoretically sound methods for the computation of probabilistic information within expert systems. We began our investigation by implementing and evaluating two previously developed exact inference algorithms, followed by the development of a hybrid algorithm that combines their relative strengths. Subsequently, we designed and implemented a new type of exact inference algorithm based on recursive decomposition. Our conclusion regarding current exact algorithms for belief-network inference is that each has its strengths and weaknesses; no one algorithm is best for all inference problems. Furthermore, our analysis of the theoretical complexity of the belief-network inference problem indicates that it is unlikely we can develop an exact algorithm that is uniformly efficient (polynomial time) across all networks and inference problems. This led us to investigate special-case and approximation algorithms, as well as methods for controlling multiple algorithms in solving a single inference problem. Our investigation indicates that moderately complex belief-network expert systems can be constructed using these current methods. Additional research is needed to understand better how to control the application of multiple algorithms to solve a single probabilistic inference task. We are continuing to explore this area of research.

The construction of complex belief networks also presents significant challenges. We have developed automated and semiautomated knowledge-acquisition techniques that show substantial promise in preliminary tests. The automated acquisition of belief networks from databases appears to be particularly promising. We believe that further exploration of automated methods for the acquisition of belief networks from databases has excellent potential to yield significant new results.

List of Publications

1. I.A. Beinlich, H.J. Suermondt, R.M. Chavez and G.F. Cooper, The ALARM monitoring system: A case study with two probabilistic inference techniques for belief networks, *Proceedings of the Second European Conference on Artificial Intelligence in Medicine* (1989) 247–256.

2. R.M. Chavez and G.F. Cooper, An empirical evaluation of a randomized algorithm for probabilistic inference, In: M. Henrion, R.D. Shachter, L.N. Kanal and J.F. Lemmer (Eds.) *Uncertainty in Artificial Intelligence* 5 (North-Holland, Amsterdam, 1990) 191–207.

3. R.M. Chavez, Hypermedia and randomized algorithms for medical expert systems, In: *Proceedings of the Symposium on Computer Applications in Medical Care* (1989) 171–177.

4. R.M. Chavez and G.F. Cooper, A randomized approximation algorithm for probabilistic inference on Bayesian belief networks, *Networks* **20** (1990) 661–685.

5. R.M. Chavez and G.F. Cooper, A randomized approximation analysis of logic sampling, *Proceedings of the Conference on Uncertainty in Artificial Intelligence* (1990) 130–135.

6. H.L. Chin and G.F. Cooper, Bayesian belief network inference using simulation, In: L.N. Kanal, T.S. Levitt and J.F. Lemmer (Eds.) *Uncertainty in Artificial Intelligence* 3 (North-Holland, Amsterdam, 1989) 129–147.

7. G.F. Cooper, An algorithm for computing probabilistic propositions, In: L.N. Kanal, T.S. Levitt and J.F. Lemmer (Eds.) *Uncertainty in Artificial Intelligence* 3 (North-Holland, Amsterdam, 1989) 3–14.

8. G.F. Cooper, The computational complexity of probabilistic inference using Bayesian belief networks, *Artificial Intelligence* **42** (1990) 393–405.

9. G.F. Cooper and E.H. Herskovits, A Bayesian method for the induction of probabilistic networks from data, *Machine Learning* (to appear).

10. G.F. Cooper, Bayesian belief-network inference using recursive decomposition (to be submitted for publication).

11. D.E. Heckerman and E.J. Horvitz, Problem formulation as the reduction of a decision model, In: *Proceedings of the Conference on Uncertainty in Artificial Intelligence* (1990) 82–89.

12. E.H. Herskovits and G.F. Cooper, Kutato: An entropy-driven system for construction of probabilistic expert systems from databases, In: *Proceedings of the Conference on Uncertainty in Artificial Intelligence* (1990) 54–62.

13. E.H. Herskovits and G.F. Cooper, Algorithms for belief network precomputation, *Methods of Information in Medicine* **30** (1991) 81–89.

14. E.J. Horvitz, G.F. Cooper and D.E. Heckerman, Reflection and action under scarce resources: Theoretical principles and empirical study, *Proceedings of the International Joint Conference on Artificial Intelligence* (1989) 1121–1127.

15. E.H. Horvitz, H.J. Suermondt and G.F. Cooper, Bounded conditioning: Flexible inference for decisions under scarce resources, In: *Proceedings of the Workshop on Uncertainty in Artificial Intelligence* (1989) 182–193.

16. B.F. Middleton, M.A. Shwe, D.E. Heckerman, M. Henrion, E.J. Horvitz, H. Lehmann and G.F. Cooper, Probabilistic diagnosis using a reformulation of the INTERNIST-1/QMR knowledge base: Evaluation of diagnostic performance, *Methods of Information in Medicine* (to appear).

17. M.A. Shwe and G.F. Cooper, An empirical analysis of likelihood-weighting simulation on a large, multiply connected medical belief network, *Computers and Biomedical Research* **24** (1991) 453–475.

18. M.A. Shwe, B.F. Middleton, D.E. Heckerman, M. Henrion, E.J. Horvitz, H. Lehmann and G.F. Cooper, Probabilistic diagnosis using a reformulation of the INTERNIST 1/QMR knowledge base: The probabilistic model and inference algorithms, *Methods of Information in Medicine* (in press).

19. H.J. Suermondt and G.F. Cooper, Probabilistic inference in multiply connected belief networks using loop cutsets, *International Journal of Approximate Reasoning* **4** (1990) 283–306.

20. H.J. Suermondt, G.F. Cooper and D.E. Heckerman, A combination of cutset conditioning with clique-tree propagation in the Pathfinder system, *Proceedings of the Conference on Uncertainty in Artificial Intelligence* (1990) 273–279.

21. H.J. Suermondt and G.F. Cooper, A combination of exact algorithms for inference on Bayesian belief networks, *International Journal of Approximate Reasoning* **5** (1991) 521–542.

22. H.J. Suermondt and G.F. Cooper, Initialization for the method of conditioning in Bayesian belief networks, *Artificial Intelligence* **50** (1991) 83–94.

23. H.J. Suermondt, Explanation of probabilistic inference in Bayesian belief networks, Report KSL-91-39, Knowledge Systems Laboratory, Medical Computer Science, Stanford University, Stanford, CA.

Dissertations completed or in progress

24. R. M. Chavez, *Architectures and Approximation Algorithms for Probabilistic Expert Systems*, doctoral dissertation in Medical Information Sciences at Stanford University (completed December 1990).

25. E.H. Herskovits, *Computer-Based Probabilistic-Network Construction*, doctoral dissertation in Medical Information Sciences at Stanford University (completed June 1991).

26. E.J. Horvitz, *Computation and Action Under Bounded Resources*, doctoral dissertation in Medical Information Sciences at Stanford University (completed December 1990).

27. H.J. Suermondt, *Explanation in Bayesian Belief Networks*, doctoral dissertation in Medical Information Sciences at Stanford University (completion is expected by March 1992).

List of Participating Scientific Personnel

Principal Investigator: Edward H. Shortliffe

Project Director: Gregory F. Cooper

Graduate Students:

R. M. Chavez — obtained a Ph.D. in Medical Information Sciences at Stanford University in December 1990.

E.H. Herskovits, — obtained a Ph.D. in Medical Information Sciences at Stanford University in June 1991.

E.J. Horvitz — obtained a Ph.D. in Medical Information Sciences at Stanford University in December 1990.

M.A. Shwe — obtained an M.S. in Medical Information Sciences at Stanford University in January 1990.

H.J. Suermondt — obtained an M.S. in Medical Information Sciences at Stanford University in June 1989, and expects to complete a Ph.D. by March 1992.

Information on Inventions

This project produced no inventions.